

## Forecasting the yield curve with the arbitrage-free dynamic Nelson–Siegel model: Brazilian evidence

João F. Caldeira<sup>a,\*</sup>, Guilherme V. Moura<sup>b</sup>, André A.P. Santos<sup>b</sup>, Fabricio Tourrucô<sup>a</sup>

<sup>a</sup> Department of Economics & PPGA, Universidade Federal do Rio Grande do Sul, Brazil

<sup>b</sup> Department of Economics, Universidade Federal de Santa Catarina, Brazil

Received 19 October 2014; received in revised form 2 May 2015; accepted 22 June 2016

Available online 25 July 2016

### Abstract

We assess the extent to which the imposition of a no-arbitrage restriction on the dynamic Nelson–Siegel model helps obtaining more accurate forecasts of the term structure. For that purpose, we provide an empirical application based on a large panel of Brazilian interest rate future contracts and test for differences in forecasting performance among alternative benchmark specifications including the random walk, vector autoregressions, and the dynamic Nelson–Siegel. We show empirically that the arbitrage-free Nelson–Siegel model is able to outperform all other benchmark models when longer forecasting horizons are taken into account.

© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of National Association of Postgraduate Centers in Economics, ANPEC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Yield curve; Arbitrage-free Nelson–Siegel model; Dynamic factor models; Kalman filter

### Resumo

Neste artigo avaliamos em que medida a imposição de uma restrição de não arbitragem na versão dinâmica do modelo de Nelson–Siegel ajuda a obter previsões mais precisas da estrutura a termo. Para isso, realizamos uma aplicação empírica envolvendo um amplo conjunto de taxas de juros de contratos de DI-futuro negociados na BM&F Bovespa. Os resultados são comparados com os modelos competidores mais amplamente usados, incluindo o *random walk*, vetores autorregressivos e o modelo dinâmico de Nelson–Siegel. Os resultados encontrados mostram evidências de que o modelo de Nelson–Siegel com condição de não arbitragem é capaz de superar os modelos benchmarks quando se consideram horizontes de previsão mais longos em todo o espectro de maturidades analisadas.

© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of National Association of Postgraduate Centers in Economics, ANPEC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Palavras chave:** Curva de juros; Modelo Nelson–Siegel livre de arbitragem; Modelos de fatores dinâmicos; Filtro de Kalman

\* Corresponding author at: Department of Economics, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 90040-000, Brazil. Fax: +55 51 33083440.

E-mail address: [emaildocaldeira@gmail.com](mailto:emaildocaldeira@gmail.com) (J.F. Caldeira).

Peer review under responsibility of National Association of Postgraduate Centers in Economics, ANPEC.

<http://dx.doi.org/10.1016/j.econ.2016.06.003>

1517-7580 © 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of National Association of Postgraduate Centers in Economics, ANPEC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

There has been growing interest in the ability to forecast the behavior of the term structure of interest rates. Such forecasts are of paramount importance for macroeconomists, financial economists, and fixed income managers since bond portfolio optimization, risk management, and pricing of financial assets and their derivatives rely heavily on interest rate forecasts. Moreover, these forecasts are widely used by financial institutions, regulators, and institutional investors to develop macroeconomic scenarios.

One of the most popular approaches to forecasting the yield curve is the dynamic version of the Nelson and Siegel (1987) model proposed by Diebold and Li (2006) (hereafter DNS). Existing evidence suggests that these specifications are remarkably well suited both to fit the term structure and to forecast its movements. Vicente and Tabak (2008) compared a Gaussian affine model with Diebold and Li model for Brazilian data and concluded that the latter model is slightly superior in terms of yield curve forecasts. Vereda et al. (2008) employ a VAR approach to forecast the term structure of interest rates and find that incorporating macro variables can improve forecasting performance, especially for longer-term forecasts. Almeida et al. (2009) obtained good forecasting results using an expanded version of the Nelson–Siegel model proposed by Svensson (1994) to accommodate additional nonlinearities in emerging market data. de Rezende and Ferreira (2013) proposed a five factor version of the Nelson–Siegel model and showed that it improves in-sample fit, however, out-of-sample results favored more parsimonious models. In Caldeira et al. (2010) the Nelson–Siegel model is cast in state-space form, and the parameters are simultaneously and efficiently estimated using the Kalman filter.<sup>1</sup>

Despite the large empirical evidence favorable to the DNS approach, one drawback is that it fails on an important theoretical dimension: it does not impose restrictions to prevent riskless arbitrage opportunities, as shown in Björk and Christensen (1999). This drawback is relevant, since many financial applications that rely on interest rate modeling such as the pricing of interest-rate-linked assets require an arbitrage-free setting. This difficulty motivated Christensen et al. (2009, 2011) to develop an arbitrage-free version of the DNS model (hereafter AFNS), thus overcoming the theoretical weakness of the original model specification.

The AFNS model of Christensen et al. (2009, 2011) has many appealing features. First, it preserves the desirable economic interpretation of the three-factor model of time-varying level, slope and curvature of the DNS specification. Second, AFNS ensures lack of arbitrage opportunities with a more parsimonious structure in comparison to general affine arbitrage-free models such as those considered in Duffie and Kan (1996) and Duffee (2002). More specifically, (Christensen et al., 2011) show that to achieve these desirable properties, one only needs to add a yield-adjustment term containing the necessary restrictions to the DNS specification.

That being said, an immediate question arises: is the no-arbitrage imposition helpful for forecasting purposes? This question is, indeed, very controversial. First, as we shall see in Section 2.2, the yield-adjustment term of the AFNS model puts no restriction on the dynamics of the yields and Joslin et al. (2011) show theoretically that no-arbitrage conditions cannot improve forecasts of the risk factors. In other words, the imposition of no-arbitrage delivers yield-adjustment terms that vary with maturity but are constant over time. This suggests that the inclusion of the yield-adjustment term is unlikely to provide forecasting gains. In particular, Duffee (2011) points out that imposition of no-arbitrage based on cross-section restrictions is irrelevant for forecasting.<sup>2</sup> However, empirical work has found predictive gains from imposing no-arbitrage. For instance, Ang and Piazzesi (2003), Favero et al. (2012) and Moench (2008) find that imposition of no-arbitrage often improves VAR forecasts, while Almeida and Vicente (2008) corroborate these results using polynomial models. Gimeno and Marqués (2009) and Christensen et al. (2011) used different data sets and found that imposition of no-arbitrage leads to substantial forecast improvements. Diebold and Rudebusch (2013) point out that, despite its time constancy, the yield-adjustment term can act as a bias correction and thus produce forecast improvements.<sup>3</sup> However, there is no clear-cut theoretical result showing that no-arbitrage restrictions improve forecasts, implying that all empirical results discussed above are data and model dependent. Thus, additional empirical

<sup>1</sup> Diebold and Rudebusch (2013) and the references therein present international evidence on the forecasting ability of DNS-type models.

<sup>2</sup> Coroneo et al. (2011) and Nyholm and Vidova-Koleva (2012) also found that the imposition of no-arbitrage adds little to forecasting accuracy.

<sup>3</sup> Moreover, Carriero and Giacomini (2011) show that the imposition of no-arbitrage restriction is important specially when an economic measure of accuracy is taken into account.

work is necessary in order to help clarify the gains (or lack thereof) stemming from the imposition of no-arbitrage in the DNS class of models estimated based on Brazilian data.

In this paper we assess the extent to which the imposition of a no-arbitrage restriction on the DNS setting brings additional forecasting gains by providing an empirical evidence based on a large data set of constant-maturity future contracts of the Brazilian Inter Bank Deposit Contract (DI1) which is equivalent to a zero-coupon bond and is highly liquid (293 million contracts worth US\$ 15 billion traded in 2010). The market for DI1 contracts is one of the most liquid interest rate markets in the world. Many banks, insurance companies, and investors use DI1 contracts as investment and hedging instruments. The data set considered in the paper contains daily observations of DI1 contracts traded on the Brazilian Mercantile and Futures Exchange (BM&F) with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42, and 48 months. We use a rolling estimation to produce out-of-sample forecasts for 1-week, 1-month, 3-month, and 6-month ahead based on the DNS and AFNS models with both uncorrelated and correlated factors. Moreover, we consider three alternative classical benchmark specifications: the random walk (RW) model, which is taken as our baseline model, the unrestricted first-order autoregressive model (AR(1)), and the first-order vector-autoregressive model (VAR(1)). We assess forecasting accuracy of a model for each of the 14 maturities by means of the root mean squared forecast error (RMSFE), and the overall performance of a given model by the trace RMSFE (TRMSFE) considered in [Hoerdahl et al. \(2006\)](#), [de Pooter et al. \(2010\)](#). Finally, we also test for the differences in forecasting performance using the test proposed in [Giacomini and White \(2006\)](#).

Our empirical evidence suggests that when a short forecasting horizon is considered (e.g. 1-week-ahead), the differences in forecasting performance among the candidate models is rather inconclusive since in very few instances the baseline RW model is outperformed. However, it is possible to see that when longer forecasting horizons are considered, the AFNS model with uncorrelated factors appears to deliver the most accurate forecasts. Therefore, the most important message from our empirical test is that the imposition of no-arbitrage is indeed helpful, but only for longer (e.g. 3-month- and 6-month-ahead) forecasting horizons. These results corroborate the evidence reported in [Christensen et al. \(2011\)](#) for US Treasuries data, as they find that the AFNS model with independent factors outperforms the DNS specifications for the 6-month- and 12-month-ahead forecast horizons. Our results corroborate the findings in [Araújo and Cajueiro \(2013\)](#), [Caldeira et al. \(2016\)](#), who shows that it is not possible to determine an individual model that consistently produces superior forecasts for all maturities and all forecasts horizons. Nevertheless, empirical results suggest that the traditional DNS model has good out-of-sample forecasting performance when compared to the RW, AR(1), and VAR(1), specially when we consider 1- and 3-month ahead horizon.

The outline of the paper is as follows. Section 2 describes the DNS and AFNS specifications adopted in this paper. Section 3 discusses the implementation details and empirical results. Finally, Section 4 concludes.

## 2. The Nelson–Siegel class of models

[Nelson and Siegel \(1987\)](#) have shown that the term structure can be surprisingly well fitted at a particular point in time by a linear combination of three smooth functions. The Nelson–Siegel model of the yield curve is given by

$$y(\tau) = \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_\tau \quad (1)$$

where  $y(\tau)$  is the zero-coupon yield with  $\tau$  months to maturity, and  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  can be interpreted as the level, slope, and curvature of the yield curve, respectively. The parameter  $\lambda$  determines the exponential decay of the  $\beta_2$  and  $\beta_3$  loadings.

### 2.1. Dynamic Nelson–Siegel model (DNS)

[Diebold and Li \(2006\)](#) show that using the static Nelson–Siegel model as basis for a dynamic factor model generates highly accurate interest rate forecasts. The dynamic Nelson–Siegel model (DNS) is given by

$$y_t(\tau) = X_{1t} + X_{2t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + X_{3t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau), \quad (2)$$

where  $X_{1t}$ ,  $X_{2t}$ , and  $X_{3t}$  can be interpreted as the time-varying level, slope, and curvature factors.

The DNS model can be interpreted as a dynamic factor model and written in state-space form. Consider the  $N \times T$  matrix of observable yields ( $\mathbf{y}_t$ ),  $t = 1, \dots, T$ , where  $\mathbf{y}_t$  is the observation vector at time  $t$ ,  $\mathbf{y}_t = (y_{it})$ ,  $i = 1, \dots, N$ . The DNS model can be represented in state-space form as

$$\mathbf{y}_t = \Gamma + B\mathbf{X}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \quad t = 1, \dots, T, \quad (3)$$

$$\mathbf{X}_t = \mu + A\mathbf{X}_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Omega), \quad t = 1, \dots, T, \quad (4)$$

where  $B$  is the  $N \times K$  matrix of factor loadings that depends on the decay parameter  $\lambda$ ,  $\mathbf{X}_t = (X_{1t}, \dots, X_{Kt})'$  is a  $K$ -dimensional vector containing the coefficients,  $\varepsilon_t$  is the  $N \times 1$  vector of disturbances with  $\Sigma$  being its  $N \times N$  diagonal covariance matrix.  $\mu$  is a  $K \times 1$  vector of constants,  $A$  is the  $K \times K$  transition matrix, and  $\Omega$  is the conditional covariance matrix of disturbance vector  $\eta_t$ , which are independent of the residuals  $\varepsilon_t \forall t$ .  $\Gamma$  is a vector of constants, which is fixed to zero in the DNS specification but will be different from zero in the AFNS specification presented below. Eqs. (3) and (4) characterize a general Gaussian linear state-space model.

Empirically, the DNS model is highly tractable and provides good fits. Theoretically, however, arbitrage opportunities cannot be prevented due to an unrestricted dynamic evolution of the yields. Indeed, as implied by Filipovic (1999), it is impossible to prevent arbitrage at bond prices in the resulting Nelson–Siegel yield curve. Christensen et al. (2011) showed how to remedy this theoretical weakness.

## 2.2. The dynamic arbitrage-free Nelson–Siegel model (AFNS)

Aiming to overcome this theoretical weakness, but hoping to maintain the good properties of the DNS model, (Christensen et al., 2011) searched for an arbitrage-free yield curve model in the general specification proposed by Duffie and Kan (1996) which has the same factor loadings as the DNS model.

The general structure of Duffie and Kan (1996) involves considering the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$  satisfying the usual conditions (see Williams, 1997).  $\mathbb{Q}$  denotes the risk-neutral measure and we will denote the real world probability measure by  $\mathbb{P}$ . The risk-neutral dynamic factors  $\mathbf{X}_t$  are assumed to follow a Markov process defined on a set  $M \subseteq \mathbb{R}^n$  that solves the stochastic differential equation (SDE)

$$d\mathbf{X}_t = \mathbf{K}^\mathbb{Q}(t) \left[ \Theta^\mathbb{Q}(t) - \mathbf{X}_t \right] dt + \Sigma(t) D(\mathbf{X}_t, t) d\mathbf{W}_t^\mathbb{Q}, \quad (5)$$

where  $\mathbf{W}^\mathbb{Q}$  is a standard Brownian Motion in  $\mathbb{R}^n$  defined according to the filtration  $(\mathcal{F}_t)_{t \geq 0}$ . Furthermore, as in Duffie and Kan (1996) one assumes that the drifts,  $\Theta^\mathbb{Q} : [0, T] \rightarrow \mathbb{R}^n$ , dynamics,  $\mathbf{K}^\mathbb{Q} : [0, T] \rightarrow \mathbb{R}^{n \times n}$ , and volatility,  $\Sigma : [0, T] \rightarrow \mathbb{R}^{n \times n}$ , are continuous functions. Finally, it is assumed that the mapping  $D : M \times [0, T] \rightarrow \mathbb{R}^{n \times n}$  has a diagonal structure with entries

$$D_{ii} = \sqrt{\gamma^i(t) + \delta^1(t)X_{1t} + \dots + \delta^n(t)X_{nt}} \quad \forall i \in \{1, \dots, n\}, \quad (6)$$

where  $\gamma^i : [0, T] \rightarrow \mathbb{R}^n$  and  $\delta^i : [0, T] \rightarrow \mathbb{R}^{n \times n}$  are continuous functions. In addition, Duffie and Kan (1996) assume the instantaneous risk-neutral rate is an affine function of the state variables

$$r_t = \rho_0(t) + \rho_1(t)' \mathbf{X}_t, \quad (7)$$

with continuous functions  $\rho_0 : [0, T] \rightarrow \mathbb{R}$  and  $\rho_1 : [0, T] \rightarrow \mathbb{R}^n$ . Under this affine formulation Duffie and Kan (1996) proved that a closed-form analytic expression for zero-coupon bond prices is attained as a linear function of the latent dynamic factors. This implies that the zero-coupon yield at time  $t$  for a bond with maturity  $T$  is given by:

$$y(t, T) = -\frac{1}{T-t} \log \left( \mathbb{E}^\mathbb{Q} \left[ \exp \left( -\int_t^T r_u du \right) \right] \right) = -\frac{1}{T-t} [\Gamma(t, T) + B(t, T)' \mathbf{X}_t] \quad (8)$$

where  $\Gamma(t, T)$  and  $B(t, T)$  satisfy the system of ordinary differential equations (ODEs)

$$\frac{d\Gamma(t, T)}{dt} = \rho_0 - B(t, T)' (\mathbf{K}^\mathbb{Q})' \Theta^\mathbb{Q} - \frac{1}{2} \sum_{i=1}^n [\Sigma' B(t, T) B(t, T)' \Sigma]_{ii} \gamma^i, \quad (9)$$

$$\frac{dB(t, T)}{dt} = \rho_1 + (\mathbf{K}^{\mathbb{Q}})' B(t, T) - \frac{1}{2} \sum_{i=1}^n [\Sigma' B(t, T) B(t, T)' \Sigma]_{ii} (\delta^i)', \quad (10)$$

with boundary conditions  $\Gamma(T, T) = 0$  and  $B(T, T) = 0$ .

Christensen et al. (2011) objective was to work under the generic affine dynamic latent factor structure of Duffie and Kan (1996) stated in (8), which rules out arbitrage opportunities. However, this general specification does not guarantee that the solution to (9)–(10) delivers factor loadings similar to the ones from the DNS model. In order to obtain the desired  $B(t, T)$ , it is necessary to impose additional restrictions in the system (9)–(10) to ensure that it has solutions that match the factor loadings of the DNS model. Specifically, Proposition 1 of Christensen et al. (2011) shows that to obtain the DNS factor loadings it is necessary to assume that  $\rho_0 = 0$ ,  $\rho_1 = (1, 1, 0)'$ ,  $\delta = 0$ ,  $\gamma = (1, \dots, 1)'$ , and to restrict the state-variable reversion rates under the  $\mathbb{Q}$  measure to obtain

$$K^{\mathbb{Q}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix},$$

which imply that the instantaneous rate is given by  $r_t = X_{1t} + X_{2t}$ , and that the latent state variables  $\mathbf{X}_t = (X_{1t}, X_{2t}, X_{3t})$  are described by the following system of stochastic differential equations under the risk-neutral measure  $\mathbb{Q}$ :

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} = -K^{\mathbb{Q}} \left[ \begin{pmatrix} \theta_1^{\mathbb{Q}} \\ \theta_2^{\mathbb{Q}} \\ \theta_3^{\mathbb{Q}} \end{pmatrix} - \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_{1t}^{\mathbb{Q}} \\ dW_{2t}^{\mathbb{Q}} \\ dW_{3t}^{\mathbb{Q}} \end{pmatrix}, \quad \lambda > 0.$$

Given these restrictions, Christensen et al. (2011) prove that the factor loadings  $B(t, T)$  that solve (9)–(10) are given by:

$$\begin{aligned} B^1(t, T) &= -(T - t), \\ B^2(t, T) &= -\frac{1 - e^{-\lambda(T-t)}}{\lambda}, \\ B^3(t, T) &= -\frac{1 - e^{-\lambda(T-t)}}{\lambda} + (T - t)e^{-\lambda(T-t)}. \end{aligned} \quad (11)$$

Substituting (11) in (8) gives:

$$y(t, T) = X_{1t} + X_{2t} \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right] + X_{3t} \left[ \frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right] - \frac{\Gamma(t, T)}{T-t}, \quad (12)$$

which differs from the DNS measurement equation given in (2) only by the additional time-independent term  $-\frac{\Gamma(t, T)}{T-t}$  that ensures the model is arbitrage-free.

Additionally, note that the drift terms  $\Theta^{\mathbb{Q}}$ , and the volatility matrix  $\Sigma$  are not in the system of ODEs for the  $B(t, T)$  functions described in (10), indicating that they do not affect the solution (11). The drift term and the volatility matrix appear only in (9), and thus need to be identified based on the yield-adjustment term  $-\frac{\Gamma(t, T)}{T-t}$ . However, as it is typical in factor models,<sup>4</sup> it is not possible to simultaneously identify the drift term affecting the unobserved factors,  $\Theta^{\mathbb{Q}}$ , and the intercept,  $-\frac{\Gamma(t, T)}{T-t}$ . Therefore, Christensen et al. (2011) fixed  $\Theta^{\mathbb{Q}} = 0$ , and showed that yield-adjustment term becomes:

$$\frac{\Gamma(t, T)}{T-t} = \frac{1}{2} \frac{1}{T-t} \sum_{i=1}^3 \int_t^T [\Sigma' B(s, T) B(s, T)' \Sigma]_{ii} ds, \quad (13)$$

<sup>4</sup> See Geweke and Zhou (1996) for a detailed discussion of identification issues in factor models.

Eq. (13) makes it clear that the volatility matrix  $\Sigma$  is only identified through the product  $\Sigma' B(s, T) B(s, T)' \Sigma$ , and knowledge of  $B(s, T)$  and  $\Gamma(t, T)$  allows only the identification of a triangular matrix  $\Sigma$  (see Christensen et al., 2011, Appendix B):

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

### 2.2.1. Independent and correlated AFNS models

We consider two versions of the DNS model and investigate the effect of the arbitrage-free restriction on their corresponding AFNS versions. The first is the uncorrelated-factor DNS model which is specified in terms of a diagonal transition matrix in (4). The correlated-factor DNS model has a full transition matrix instead. In both DNS models the measurement equation is the same.

The corresponding AFNS models are formulated in continuous time and the relation between the dynamics under measures  $\mathbb{Q}$  and  $\mathbb{P}$  is given by a change of measure that preserves the affine dynamic structure (see Christensen et al., 2011 for full details). This is a very convenient property since it guarantees that identical ( $\mathbb{P}$ -measure) models can be estimated.

Using the solution already presented in (12), yields must be related to the state variables by

$$\mathbf{y}_t = \Gamma + B X_t + \varepsilon_t. \quad (14)$$

Comparing equations (2) and (12), we notice that the matrix  $B$  is identical for DNS and AFNS models. The only difference is the added vector  $\Gamma$  containing the yield-adjustment terms in the AFNS models. Since AFNS is a continuous-time model, the dynamics is discretized to allow comparison with the discrete-time DNS models. The estimation is outlined in the next section.

For the uncorrelated-factor AFNS model the dynamics of the state variables under the  $\mathbb{P}$ -measure is

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{\mathbb{P}} & 0 & 0 \\ 0 & \kappa_{22}^{\mathbb{P}} & 0 \\ 0 & 0 & \kappa_{33}^{\mathbb{P}} \end{pmatrix} \left[ \begin{pmatrix} \theta_1^{\mathbb{P}} \\ \theta_2^{\mathbb{P}} \\ \theta_3^{\mathbb{P}} \end{pmatrix} - \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_{1t}^{\mathbb{P}} \\ dW_{2t}^{\mathbb{P}} \\ dW_{3t}^{\mathbb{P}} \end{pmatrix}. \quad (15)$$

In the correlated-factor AFNS model, the three shocks may be correlated, and there may be full interaction among the factors as they adjust to the steady state

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \\ dX_{3t} \end{pmatrix} = \begin{pmatrix} \kappa_{11}^{\mathbb{P}} & \kappa_{12}^{\mathbb{P}} & \kappa_{13}^{\mathbb{P}} \\ \kappa_{21}^{\mathbb{P}} & \kappa_{22}^{\mathbb{P}} & \kappa_{23}^{\mathbb{P}} \\ \kappa_{31}^{\mathbb{P}} & \kappa_{32}^{\mathbb{P}} & \kappa_{33}^{\mathbb{P}} \end{pmatrix} \left[ \begin{pmatrix} \theta_1^{\mathbb{P}} \\ \theta_2^{\mathbb{P}} \\ \theta_3^{\mathbb{P}} \end{pmatrix} - \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_{1t}^{\mathbb{P}} \\ dW_{2t}^{\mathbb{P}} \\ dW_{3t}^{\mathbb{P}} \end{pmatrix}. \quad (16)$$

This is the most flexible version of the AFNS models, where all parameters are identified. From these specifications we see the striking resemblance between DNS and AFNS models.

### 2.2.2. An exact expression for the covariance of the continuous-time AFNS model

In this paper, the models are estimated using a maximum likelihood estimation method based on the Kalman filter discussed in Section 2.3. The AFNS model can be formulated in a state-space form as follows. Departing from the continuous-time formulation of the AFNS model, the conditional mean vector and the conditional covariance matrix are

$$\mathbb{E}^{\mathbb{P}} [\mathbf{X}_T | \mathcal{F}_t] = \left[ I - \exp(-\mathbf{K}^{\mathbb{P}} \Delta t) \right] \Theta^{\mathbb{P}} + \exp(-\mathbf{K}^{\mathbb{P}} \Delta t) X_t \quad (17)$$

$$\mathbb{V}^{\mathbb{P}} [\mathbf{X}_T | \mathcal{F}_t] = \int_0^{\Delta t} \exp(-\mathbf{K}^{\mathbb{P}} s) \Sigma \Sigma' \exp(-\mathbf{K}^{\mathbb{P}})' s ds \quad (18)$$

To estimate the AFNS models proposed by [Christensen et al. \(2011\)](#) we must compute the conditional covariance matrix

$$\mathbb{V}^{\mathbb{P}}[\mathbf{X}_t | \mathcal{F}_{t-1}] = \int_0^{\Delta t} \exp(-\mathbf{K}^{\mathbb{P}} s) \Sigma \Sigma' \exp(-(\mathbf{K}^{\mathbb{P}})' s) ds \quad (19)$$

of discrete observations. Since the estimation is an intense computational process, we need to provide fast intermediate calculations. One approach is to approximate the integral and the matrix exponential. Another approach uses the diagonalization of  $\mathbf{K}^{\mathbb{P}}$  to calculate the integral exactly. Due to the reduced size of the matrix  $\mathbf{K}^{\mathbb{P}}$ , the latter is significantly faster than the former.

The AFNS state transition is

$$\mathbf{X}_t = \left( I - \exp(-\mathbf{K}^{\mathbb{P}} \Delta t) \right) \Theta^{\mathbb{P}} + \exp(-\mathbf{K}^{\mathbb{P}} \Delta t) \mathbf{X}_{t-1} + \eta_t,$$

where  $\Delta t$  is the time between the observations at  $t$  and  $t - 1$ , with measurement equation

$$\mathbf{y}_t = \Gamma + B\mathbf{X}_t + \varepsilon_t,$$

and error structure

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix} \right],$$

where  $H$  is diagonal,  $Q = \mathbb{V}^{\mathbb{P}}[X_t | \mathcal{F}_{t-1}]$  and the transition and measurement errors are assumed orthogonal to the initial state.

The computation of  $Q$  is straightforward if we use the diagonalization of  $\mathbf{K}^{\mathbb{P}}$

$$\mathbf{K}^{\mathbb{P}} = V \Lambda V^{-1}, \quad (20)$$

where  $V$  contains the eigenvectors of  $\mathbf{K}^{\mathbb{P}}$ , and  $\Lambda$  is a diagonal matrix containing the eigenvalues  $(\lambda_i)$  of  $\mathbf{K}^{\mathbb{P}}$ . We refer the reader to [Lang \(1987\)](#) and [Golub and Van Loan \(1996\)](#) for details on linear algebra theory and on the computational aspects of linear algebra, respectively.

Substituting (20) in (19), using

$$\exp(-\mathbf{K}^{\mathbb{P}} s) = V \exp(-\Lambda s) V^{-1},$$

and similarly

$$\exp(-(\mathbf{K}^{\mathbb{P}})' s) = (V^{-1})' \exp(-\Lambda s) V',$$

we obtain

$$Q = V \left( \int_0^{\Delta t} \exp(-\Lambda s) \Omega \exp(-\Lambda s) ds \right) V',$$

where  $\Omega = (\omega_{ij})_{n \times n} = V^{-1} \Sigma \Sigma' (V^{-1})'$ . Since the exponential of a diagonal matrix with entries  $-\lambda_i s$  is a diagonal matrix with entries  $e^{-\lambda_i s}$ , each term of the matrix under the integral is  $(\omega_{ij} e^{-(\lambda_i + \lambda_j)s})_{n \times n}$ . Integration yields an expression which only involves matrix multiplications

$$Q = V \left( \frac{\omega_{ij}}{\lambda_i + \lambda_j} (1 - e^{-(\lambda_i + \lambda_j)\Delta t}) \right)_{n \times n} V'. \quad (21)$$

Stationarity of the system under the  $\mathbb{P}$ -measure is ensured if the real parts of all eigenvalues of  $\mathbf{K}^{\mathbb{P}}$  are positive, and this condition is imposed in all estimations. For this reason, we can start the Kalman filter at the unconditional mean,  $\mathbf{X}_0 = \Theta^{\mathbb{P}}$ , and covariance matrix,  $\Sigma_0$ . In particular, the unconditional variance  $\Sigma_0 = \int_0^{\infty} \exp(-\mathbf{K}^{\mathbb{P}} s) \Sigma \Sigma' \exp(-(\mathbf{K}^{\mathbb{P}})' s) ds$



used in the initialization of the filter is easily obtained from the above expression. Assuming  $\mathbf{K}^{\mathbb{P}}$  has eigenvalues with positive real parts, the integral converges to

$$\Sigma_0 = V \left( \frac{\omega_{ij}}{\lambda_i + \lambda_j} \right)_{n \times n} V'. \quad (22)$$

### 2.3. Estimation of the DNS and AFNS models

Given the state space formulation of the dynamic factor model presented in (3) and (4), the Kalman filter can be used to obtain the likelihood function via the prediction error decomposition, as well as filtered estimates of the states and of their covariance matrices. However, the computational burden associated with the Kalman filter recursions depends crucially on the dimension of both the state and observation vectors. Moreover, in yield curve models the dimension of the observation vector ( $N \times 1$ ) is often much larger than that of the state vector ( $K \times 1$ ). In these circumstances, [Jungbacker and Koopman \(2014\)](#) have shown that significant computational gains can be achieved by a simple transformation. First, define the  $N \times N$  and the  $K \times N$  matrices:

$$J = \begin{bmatrix} J^L \\ J^H \end{bmatrix}, \quad J^L = C \Lambda(\lambda)' \Sigma^{-1},$$

respectively, where  $C$  can be any  $K \times K$  invertible matrix, and  $J^H$  is chosen to guarantee that  $J$  is full rank. Selecting  $C = (\Lambda(\lambda)' \Sigma^{-1} \Lambda(\lambda))^{-1}$  implies:

$$J(y_t - \Gamma) = \begin{bmatrix} J^L(y_t - \Gamma) \\ J^H(y_t - \Gamma) \end{bmatrix} = \begin{bmatrix} y_t^L \\ y_t^H \end{bmatrix} = \begin{bmatrix} f_t \\ 0 \end{bmatrix} + \begin{bmatrix} A^L \varepsilon_t \\ A^H \varepsilon_t \end{bmatrix}, \quad \begin{bmatrix} A^L \varepsilon_t \\ A^H \varepsilon_t \end{bmatrix} \sim N \left( 0, \begin{bmatrix} C & 0 \\ 0 & \Sigma^H \end{bmatrix} \right).$$

The law of motion of the factors in (4) is not affected by the transformation. Note that  $y_t^H$  is neither dependent on  $f_t$ , nor correlated with  $y_t^L$  and, therefore, does not need to be considered for the estimation of the factors. This implies that the Kalman filter only needs to be applied to the low dimensional subvector  $y_t^L$  for signal extraction, generating large computational gains when  $N \gg K$  (see Table 1 of [Jungbacker and Koopman, 2014](#)).

Denote  $l(y)$  the log-likelihood function of the untransformed model in (3) and (4), where  $y = (y_1', \dots, y_T')'$ . Evaluation of  $l(y)$  can also take advantage of the transformations presented above. [Jungbacker and Koopman \(2014\)](#) show that the log-likelihood of the untransformed model can be represented as

$$l(y) = c + l(y^L) - \frac{T}{2} \log \frac{|\Sigma|}{|C|} - \frac{1}{2} \sum_{t=1}^T e_t' \Sigma^{-1} e_t, \quad (23)$$

where  $c$  is a constant independent of both  $y$  and the parameters,  $l(y^L)$  is the log-likelihood function of the reduced system, and  $e_t = y_t - \Gamma - \Lambda(\lambda)f_t$ . Note that computation of matrix  $J^H$  is not required at any point, as proved in Lemma 2 of [Jungbacker and Koopman \(2014\)](#).

## 3. Empirical analysis

### 3.1. Data

The data set analyzed consists of Brazilian Interbank Deposit Futures Contract (DI1).<sup>5</sup> The market for DI1 is one of the largest fixed-income markets among emerging economies. We use DI1 daily closing yields of available contracts. Since not all maturities are observed on a daily basis, we interpolate the available data using cubic splines to produce yield curves with fixed maturities. The chosen fixed maturities are: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42, and 48 months.

<sup>5</sup> BM&FBOVESPA is the entity that offers the DI1 contract and determines the number of maturities with authorized contracts.



Table 1  
Descriptive statistics for the term structure of interest rates.

| Maturity | $\tau$           | Mean  | Std Dev | Min   | Max   | Skew   | Kurt  | Acf             |                 |                  | Pacf              |                   |
|----------|------------------|-------|---------|-------|-------|--------|-------|-----------------|-----------------|------------------|-------------------|-------------------|
|          |                  |       |         |       |       |        |       | $\hat{\rho}(1)$ | $\hat{\rho}(5)$ | $\hat{\rho}(21)$ | $\hat{\alpha}(2)$ | $\hat{\alpha}(5)$ |
| Month    | 1                | 10.66 | 1.75    | 6.97  | 14.13 | −0.337 | 2.219 | 0.998           | 0.987           | 0.936            | −0.016            | −0.012            |
|          | 3                | 10.66 | 1.79    | 7.02  | 14.52 | −0.286 | 2.296 | 0.998           | 0.989           | 0.940            | −0.017            | −0.025            |
|          | 6                | 10.73 | 1.86    | 6.91  | 15.32 | −0.201 | 2.472 | 0.998           | 0.989           | 0.939            | −0.012            | −0.023            |
|          | 9                | 10.85 | 1.90    | 6.86  | 16.04 | −0.132 | 2.650 | 0.998           | 0.988           | 0.934            | 0.004             | −0.010            |
|          | 12               | 11.00 | 1.92    | 6.87  | 16.40 | −0.091 | 2.784 | 0.997           | 0.987           | 0.929            | 0.019             | 0.000             |
|          | 15               | 11.16 | 1.92    | 6.90  | 16.91 | −0.053 | 2.907 | 0.997           | 0.985           | 0.923            | 0.024             | 0.003             |
|          | 18               | 11.31 | 1.88    | 7.00  | 17.12 | −0.046 | 3.002 | 0.997           | 0.984           | 0.917            | 0.020             | 0.002             |
|          | 21               | 11.42 | 1.85    | 7.14  | 17.26 | −0.042 | 3.118 | 0.997           | 0.983           | 0.912            | 0.023             | −0.002            |
|          | 24               | 11.52 | 1.81    | 7.32  | 17.44 | −0.032 | 3.243 | 0.996           | 0.982           | 0.907            | 0.020             | −0.001            |
|          | 27               | 11.60 | 1.77    | 7.49  | 17.62 | −0.004 | 3.375 | 0.996           | 0.981           | 0.901            | 0.015             | −0.002            |
|          | 30               | 11.66 | 1.74    | 7.65  | 17.78 | 0.028  | 3.504 | 0.996           | 0.980           | 0.895            | 0.010             | −0.006            |
|          | 36               | 11.75 | 1.66    | 7.91  | 17.83 | 0.080  | 3.731 | 0.995           | 0.978           | 0.887            | 0.015             | −0.013            |
|          | 42               | 11.83 | 1.59    | 8.13  | 17.93 | 0.104  | 3.937 | 0.995           | 0.976           | 0.880            | 0.017             | −0.006            |
|          | 48               | 11.89 | 1.56    | 8.29  | 18.00 | 0.181  | 4.156 | 0.995           | 0.975           | 0.873            | 0.018             | −0.011            |
|          | <b>Level</b>     | 12.41 | 1.42    | 9.34  | 18.99 | 0.673  | 5.185 | 0.992           | 0.957           | 0.800            | −0.064            | −0.014            |
|          | <b>Slope</b>     | −1.92 | 1.95    | −6.37 | 2.84  | −0.267 | 2.814 | 0.995           | 0.974           | 0.866            | −0.039            | −0.036            |
|          | <b>Curvature</b> | −1.00 | 3.94    | −9.48 | 8.60  | 0.120  | 2.192 | 0.997           | 0.978           | 0.903            | −0.160            | −0.011            |

The Table reports summary statistics for DI1 yields over the sample period January 2007–December 2012 (1488 daily observations). We examine daily data, constructed using cubic splines. Maturity is measured in months. We show, for each maturity, mean, standard deviation, minimum, maximum and a selection of autocorrelation (Acf,  $\hat{\rho}(1)$ ,  $\hat{\rho}(5)$ , and  $\hat{\rho}(21)$ , respectively) and partial autocorrelation (Pacf,  $\hat{\alpha}(2)$  and  $\hat{\alpha}(5)$ ) coefficients.

The DI1 contract with maturity  $\tau$  is a zero-coupon future contract in which the underlying asset is the DI interest rate accrued on a daily basis, capitalized between trading days  $t$  and  $\tau$ .<sup>6</sup> The value of contract is set by its value at maturity, R\$ 100, 000.00, discounted according to the accrued interest rate negotiated between seller and buyer. In 2010 the DI1 market traded a total of 293 million contracts corresponding to US\$ 15 billion. The DI1 contract is very similar to the zero-coupon bond, except for the daily payment of margin adjustments. The data set contains liquid maturities from January 2006 to December 2012, with a total of  $T=1488$  daily observations. The data source is the Brazilian Mercantile and Futures Exchange (BM&F). Similar but monthly datasets have been considered by Araújo and Cajueiro (2013).

Table 1 reports descriptive statistics for the Brazilian interest rate yield curve based on the DI1 market. For each time series we report the mean, standard deviation, minimum, maximum and the lag-1 sample autocorrelation. The summary statistics confirm some common stylized facts to yield curve data: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, and autocorrelations are very high.

Fig. 1 displays a three-dimensional plot of the data set and illustrates how yield levels and spreads vary substantially throughout the sample. The plot also suggests the presence of an underlying factor structure. Although the yield series vary substantially over time for each maturity, a strong common pattern in the 14 series is apparent for most of them: the yield curve is an upward-sloping function of maturity. For example, the last year of the sample is characterized by rising interest rates, especially for the shorter maturities, which respond faster to the contractionary monetary policy implemented by the Brazilian Central Bank in the first half of 2010. As we can see in Fig. 2, 3-, 6-, and 12-month yields follow very closely the SELIC rate, which is the monetary policy instrument in Brazil. It is clear from Fig. 1

<sup>6</sup> The DI rate is the average daily rate of Brazilian Interbank Deposits (borrowing/lending), calculated by the Clearinghouse for Custody and Settlements (CETIP) for all business days. The DI rate, which is published on a daily basis, is expressed in annually compounded terms, based on 252 business days. When buying a DI1 contract at the  $DI^*$  rate at time  $t$  and keeping it until maturity  $\tau$ , the gain or loss is given by:

$$100.000 \left( \frac{\prod_{i=1}^{\zeta(t,\tau)} (1 + y_i)^{\frac{1}{252}}}{(1 + DI^*)^{\frac{\zeta(t,\tau)}{252}}} - 1 \right),$$

where  $y_i$  denotes the DI rate,  $(i - 1)$  days after the trading day. The function  $\zeta(t, \tau)$  represents the number of working days between  $t$  and  $\tau$ .

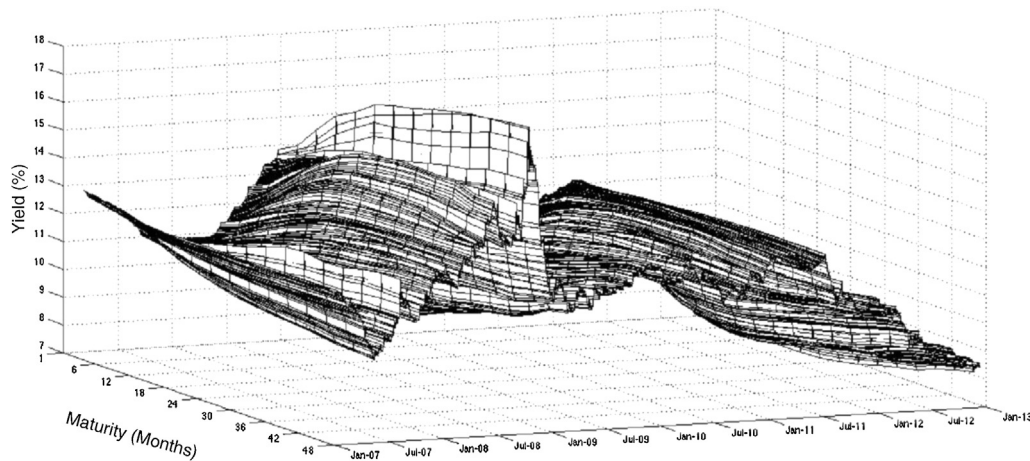


Fig. 1. Evolution of the yield curve. Evolution of the term structure of interest rates (based on DI1 contracts) for the 2006:01–2012:12 period. The sample consists of the daily yields for maturities of 1, 3, 4, 6, 9, 12, 15, 18, 24, 27, 30, 36, 42, and 48 months.

that not only the term structure level fluctuates over time but also its slope and curvature. The curve takes on various forms ranging from nearly flat to (inverted) S-type shapes.

### 3.2. Implementation details

The forecasting exercise is implemented in pseudo-real time, i.e. we never use information which is not available at the time the forecast is made. We use a rolling estimation window of 500 daily observations (2 years). We obtain forecasts for 1-week, 1-month, 3-months, and 6-months ahead. The choice of a rolling scheme is suggested by two reasons: first, it is a natural way to alleviate problems coming from parameter instability and structural breaks, see e.g. [Pesaran et al. \(2011\)](#). Second, having a fixed number of observations used to compute the forecasts results in time

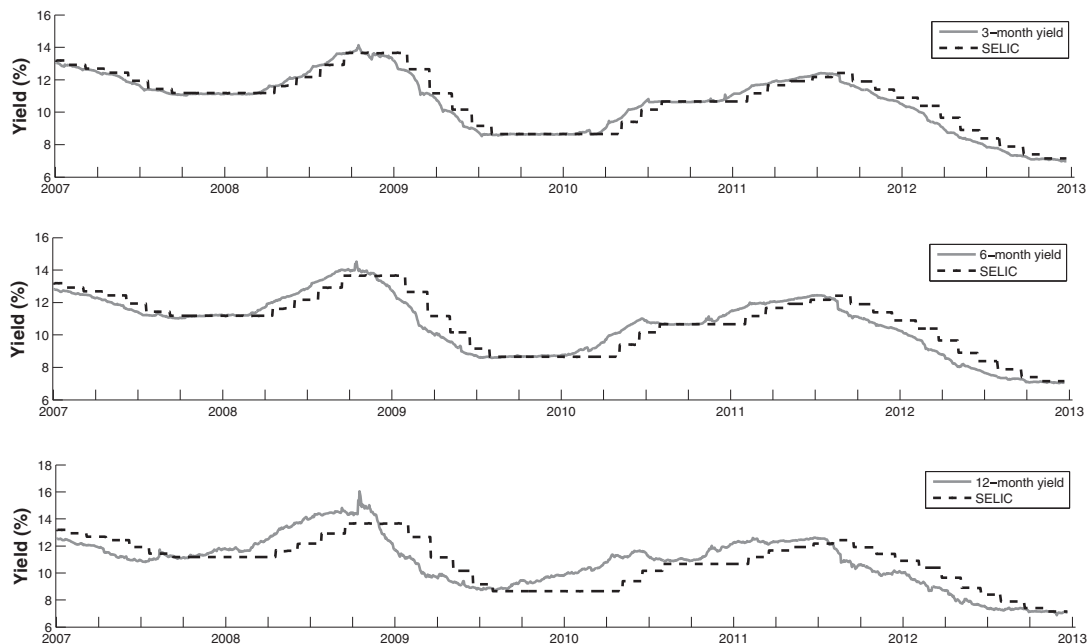


Fig. 2. US term structure dynamics over time. This figure plots the evolution of the short rates (3-, 6-, and 12-month maturities) and the realized SELIC rate, that is a natural measure of Brazilian monetary policy, for the 2006:01–2012:12 period.

series of forecast errors that can be tested for accuracy by means of the [Giacomini and White \(2006\)](#) test. Moreover, we use iterated forecasts instead of direct forecasts for the multi-period ahead predictions. [Marcellino et al. \(2005\)](#) argue that iterated forecasts are more efficient when the model is correctly specified.

In order to evaluate out-of-sample forecasts, we compute popular error metrics. Given a sample of  $\mathcal{M}$  out-of-sample forecasts for an  $h$ -period-ahead forecast horizon, we compute the root mean squared forecast error (RMSFE) for maturity  $\tau_i$  and for model  $m$  as follows:

$$\text{RMSFE}_m(\tau_i) = \sqrt{\frac{1}{\mathcal{M}} \sum_{t=1}^{\mathcal{M}} (\hat{y}_{t+h|t,m}(\tau_i) - y_{t+h}(\tau_i))^2} \quad (24)$$

where  $y_{t+h}(\tau_i)$  is the yield for the maturity  $\tau_i$  observed at time  $t+h$ , and  $\hat{y}_{t+h|t,m}(\tau_i)$  is the corresponding forecasting made at time  $t$ .

Following [Hoerdahl et al. \(2006\)](#), [de Pooter et al. \(2010\)](#), we also summarize the overall forecasting performance of each model for all maturities by computing the trace root mean squared forecast error (TRMSFE). Thus, for each forecast horizon, we compute the trace of the covariance matrix of the forecast errors across all  $N$  maturities. Hence, lower TRMSFE indicate more accurate forecasts. The TRMSFE can be computed as

$$\text{TRMSFE}_m(\tau_i) = \sqrt{\frac{1}{N} \frac{1}{\mathcal{M}} \sum_{i=1}^N \sum_{t=1}^{\mathcal{M}} (\hat{y}_{t+h|t}(\tau_i) - y_{t+h}(\tau_i))^2}, \quad (25)$$

which allows to summarize the forecasting performance of a model for all maturities in a single number.

Finally, in order to assess the statistical significance of these forecasting differences, we use the test proposed by [Giacomini and White \(2006\)](#). The [Giacomini and White \(2006\)](#) (GW) test is a conditional forecasting ability test constructed under the assumption that forecasts are generated using a moving data window. This is a test of equal forecasting accuracy and as such can handle forecasts based on both nested and non-nested models, regardless from the estimation procedures used in the derivation of the forecasts. The test is based on the loss differential  $d_{m,t} = (e_{rw,t})^2 - (e_{m,t})^2$ , where  $e_{m,t}$  is the forecast error of model  $m$  at time  $t$ . We assume that the loss function is quadratic but it can be replaced by other loss functions depending on the forecast goal. The null hypothesis of equal forecasting accuracy can be written as

$$\mathbb{H}_0 : E[d_{m,t+h} | \delta_{m,t}] = 0, \quad (26)$$

where  $\delta_{m,t}$  is a  $p \times 1$  vector of test functions or instruments and  $h$  is the forecast horizon. If a constant is used as instrument, the test can be interpreted as an unconditional test of equal forecasting accuracy. The GW test statistic  $GW_{m,t}$  can be computed as the Wald statistic:

$$GW_{m,n} = n \left( n^{-1} \sum_{t=\omega+1}^{n-h} \delta_{m,t} d_{m,t+h} \right)' \hat{\Omega}_n^{-1} \left( n^{-1} \sum_{t=\omega+1}^{n-h} \delta_{m,t} d_{m,t+h} \right) \xrightarrow{d} \chi_{\dim(\delta)}^2 \quad (27)$$

where  $\hat{\Omega}_n$  is a consistent HAC estimator for the asymptotic variance of  $\delta_{m,t} d_{m,t+h}$ , and  $n = (T - \omega)$  the number of out-of-sample observations. Under the null hypothesis given in (26), the test statistic  $GW_{i,t}$  is asymptotically distributed as  $\chi_p^2$ . Stars in [Tables 2 and 3](#) denote significant differences in out-of-sample model performance relative to RW at the 1, 5, and 10 percent levels.

### 3.2.1. Benchmark models

**3.2.1.1. Random walk model.** The main benchmark model adopted in the paper is random walk (RW), whose  $t+h$ -step-ahead forecasts for an yield of maturity  $\tau$  are given by:

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \quad \varepsilon_t(\tau_i) \sim \mathcal{N}(0, \sigma^2(\tau_i)). \quad (28)$$

In RW, a  $h$ -step-ahead forecast for yield  $\hat{y}_{t+h}(\tau_i)$  is simply equal to the most recently observed value  $y_t(\tau_i)$ . This model is a good benchmark for judging the relative prediction power of other models, since yields are usually nonstationary or nearly nonstationary. Thus, in practice, it is difficult to beat RW in terms of out-of-sample forecasting accuracy. Many other studies that consider interest rate forecasting have shown that consistently outperforming the random walk

Table 2  
Out-of-sample yield forecasts: [T] RMSFEs.

| Maturity | 1-week-ahead forecasts |              |                |                   |                    |                    |                     | 1-month-ahead forecasts |              |                |                   |                    |                    |                     |
|----------|------------------------|--------------|----------------|-------------------|--------------------|--------------------|---------------------|-------------------------|--------------|----------------|-------------------|--------------------|--------------------|---------------------|
|          | RW                     | AR(1)        | VAR(1)         | DNS <sub>AR</sub> | DNS <sub>VAR</sub> | AFNS <sub>AR</sub> | AFNS <sub>VAR</sub> | RW                      | AR(1)        | VAR(1)         | DNS <sub>AR</sub> | DNS <sub>VAR</sub> | AFNS <sub>AR</sub> | AFNS <sub>VAR</sub> |
| 3        | 0.120                  | <b>0.986</b> | <b>0.764*</b>  | <b>0.990**</b>    | 2.211              | 1.792              | 1.494               | 0.409                   | <b>0.994</b> | <b>0.573*</b>  | <b>0.995**</b>    | 1.113              | 1.192              | <b>0.576*</b>       |
| 6        | 0.141                  | <b>0.994</b> | <b>0.928**</b> | 1.384             | <b>0.961**</b>     | 1.149              | <b>0.858*</b>       | 0.430                   | <b>0.997</b> | <b>0.749**</b> | 1.008             | <b>0.960**</b>     | 1.162              | <b>0.693*</b>       |
| 9        | 0.154                  | <b>0.999</b> | 1.015          | 1.430             | 1.120              | 1.222              | 1.001               | 0.442                   | 1.001        | <b>0.858*</b>  | 1.017             | <b>0.955**</b>     | 1.164              | <b>0.845*</b>       |
| 12       | 0.164                  | <b>1.003</b> | 1.062          | 1.269             | 1.152              | 1.194              | 1.033               | 0.445                   | 1.006        | <b>0.963**</b> | 1.021             | <b>0.964**</b>     | 1.166              | <b>0.956**</b>      |
| 15       | 0.174                  | <b>1.006</b> | 1.105          | 1.104             | 1.090              | 1.126              | 1.031               | 0.455                   | 1.012        | 1.040          | 1.011             | <b>0.958**</b>     | 1.146              | 1.021               |
| 18       | 0.181                  | <b>1.008</b> | 1.130          | 1.023             | 1.045              | 1.083              | 1.033               | 0.459                   | 1.017        | 1.101          | 1.006             | <b>0.954**</b>     | 1.130              | 1.064               |
| 21       | 0.183                  | 1.011        | 1.152          | <b>1.005</b>      | 1.045              | 1.081              | 1.052               | 0.458                   | 1.023        | 1.139          | 1.002             | <b>0.953**</b>     | 1.128              | 1.101               |
| 24       | 0.184                  | 1.013        | 1.162          | <b>0.993</b>      | 1.067              | 1.079              | 1.032               | 0.453                   | 1.029        | 1.162          | <b>0.990</b>      | <b>0.948**</b>     | 1.128              | 1.104               |
| 27       | 0.187                  | 1.013        | 1.159          | <b>0.988</b>      | 1.083              | 1.074              | 1.030               | 0.453                   | 1.031        | 1.158          | <b>0.977*</b>     | <b>0.942**</b>     | 1.120              | 1.100               |
| 30       | 0.188                  | 1.014        | 1.160          | <b>0.993</b>      | 1.060              | 1.080              | 1.053               | 0.452                   | 1.036        | 1.159          | <b>0.980***</b>   | <b>0.944**</b>     | 1.119              | 1.109               |
| 36       | 0.188                  | 1.016        | 1.162          | <b>1.004</b>      | 1.035              | 1.085              | 1.072               | 0.437                   | 1.045        | 1.198          | 1.005             | <b>0.969***</b>    | 1.134              | 1.129               |
| 42       | 0.189                  | 1.018        | 1.169          | <b>1.010</b>      | 1.040              | 1.053              | 1.039               | 0.420                   | 1.056        | 1.230          | 1.009             | <b>0.978***</b>    | 1.125              | 1.107               |
| 48       | 0.192                  | <b>1.019</b> | 1.166          | 1.044             | 1.058              | 1.042              | 1.054               | 0.417                   | 1.061        | 1.230          | 1.011             | <b>0.983</b>       | 1.092              | 1.077               |
| T-RMSFE  | 0.170                  | 1.010        | 1.116          | 1.086             | 1.127              | 1.132              | 1.052               | 0.440                   | 1.024        | 1.065          | 1.002             | <b>0.969</b>       | 1.139              | 1.011               |

*Note:* This table summarizes Relative Mean Squared Forecast Errors and Trace Root Mean Squared Forecast Error (TRMPSE) relative to the random walk obtained by using each of the competing models, for the horizons 1-week, 1-, 3-, and 6-month-ahead. The first column in the table reports the value of (T) RMSFE (expressed in basis points) for the Random Walk model (RW), while all other lines reports statistics relative to the RW. The following model abbreviations are used in the table: RW stands for the random walk, (V)AR for the first-order (vector) autoregressive model, DNS for the one-step dynamic Nelson–Siegel model with a (V)AR specification for the factors, AFNS refers to the one-step arbitrage-free Nelson Siegel model with a (V)AR specification for the factors. Numbers smaller than one (shown in bold) indicate that models outperform the random walk, whereas numbers larger than one indicate underperformance. The gray box indicates outperformance in the maturity. The stars on the right of the cell entries signal the level at which the [Giacomini and White \(2006\)](#) test rejects the null of equal forecasting accuracy (\*, \*\*, and \*\*\* mean respectively rejection at 1%, 5%, and 10% level).

Table 3  
Out-of-sample yield forecasts: [T] RMSFEs.

| Maturity | 3-month-ahead forecasts |       |                |                   |                    |                    |                     | 6-month-ahead forecasts |       |        |                   |                    |                    |                     |
|----------|-------------------------|-------|----------------|-------------------|--------------------|--------------------|---------------------|-------------------------|-------|--------|-------------------|--------------------|--------------------|---------------------|
|          | RW                      | AR(1) | VAR(1)         | DNS <sub>AR</sub> | DNS <sub>VAR</sub> | AFNS <sub>AR</sub> | AFNS <sub>VAR</sub> | RW                      | AR(1) | VAR(1) | DNS <sub>AR</sub> | DNS <sub>VAR</sub> | AFNS <sub>AR</sub> | AFNS <sub>VAR</sub> |
| 3        | 1.134                   | 1.077 | <b>0.727*</b>  | <b>0.998***</b>   | 1.020              | <b>0.970***</b>    | <b>0.546*</b>       | 2.023                   | 1.252 | 1.117  | <b>0.999***</b>   | 1.113              | <b>0.758*</b>      | <b>0.583*</b>       |
| 6        | 1.192                   | 1.085 | <b>0.950**</b> | <b>0.980***</b>   | <b>0.981***</b>    | <b>0.910***</b>    | <b>0.705*</b>       | 2.092                   | 1.325 | 1.380  | <b>0.984</b>      | <b>0.960**</b>     | <b>0.710*</b>      | <b>0.697*</b>       |
| 9        | 1.234                   | 1.078 | 1.119          | <b>0.981</b>      | <b>0.967**</b>     | <b>0.881**</b>     | <b>0.806*</b>       | 2.122                   | 1.289 | 1.609  | <b>0.987</b>      | <b>0.955***</b>    | <b>0.699*</b>      | <b>0.772*</b>       |
| 12       | 1.248                   | 1.071 | 1.254          | <b>0.991</b>      | <b>0.967*</b>      | <b>0.873**</b>     | <b>0.875*</b>       | 2.110                   | 1.233 | 1.798  | <b>0.993</b>      | <b>0.964***</b>    | <b>0.703*</b>      | <b>0.825*</b>       |
| 15       | 1.252                   | 1.067 | 1.358          | <b>0.996</b>      | <b>0.966***</b>    | <b>0.869**</b>     | <b>0.919**</b>      | 2.074                   | 1.196 | 1.969  | <b>0.994</b>      | <b>0.958</b>       | <b>0.709*</b>      | <b>0.859*</b>       |
| 18       | 1.236                   | 1.066 | 1.437          | 1.003             | <b>0.969**</b>     | <b>0.871**</b>     | <b>0.953**</b>      | 2.004                   | 1.163 | 2.114  | <b>0.999</b>      | <b>0.954***</b>    | <b>0.721*</b>      | <b>0.892*</b>       |
| 21       | 1.221                   | 1.071 | 1.495          | 1.001             | <b>0.966***</b>    | <b>0.867**</b>     | <b>0.975***</b>     | 1.930                   | 1.144 | 2.252  | 1.000             | <b>0.953***</b>    | <b>0.733*</b>      | <b>0.921**</b>      |
| 24       | 1.200                   | 1.085 | 1.540          | <b>0.997</b>      | <b>0.962**</b>     | <b>0.867**</b>     | <b>0.992***</b>     | 1.856                   | 1.154 | 2.377  | <b>0.999</b>      | <b>0.948</b>       | <b>0.744*</b>      | <b>0.947**</b>      |
| 27       | 1.196                   | 1.105 | 1.566          | <b>0.985**</b>    | <b>0.952**</b>     | <b>0.859**</b>     | <b>0.996***</b>     | 1.806                   | 1.182 | 2.485  | <b>0.992**</b>    | <b>0.942***</b>    | <b>0.749*</b>      | <b>0.936*</b>       |
| 30       | 1.190                   | 1.124 | 1.597          | <b>0.985***</b>   | <b>0.952**</b>     | <b>0.857**</b>     | 1.006               | 1.760                   | 1.206 | 2.587  | <b>0.990**</b>    | <b>0.944**</b>     | <b>0.755*</b>      | <b>0.928*</b>       |
| 36       | 1.154                   | 1.144 | 1.669          | 1.001             | <b>0.969**</b>     | <b>0.862**</b>     | 1.013               | 1.641                   | 1.233 | 2.797  | <b>0.999</b>      | <b>0.969</b>       | <b>0.774*</b>      | <b>0.919**</b>      |
| 42       | 1.106                   | 1.176 | 1.709          | 1.007             | <b>0.976**</b>     | <b>0.859**</b>     | <b>0.981**</b>      | 1.534                   | 1.284 | 2.949  | 1.005             | <b>0.978</b>       | <b>0.785*</b>      | <b>0.905**</b>      |
| 48       | 1.084                   | 1.205 | 1.724          | 1.003             | <b>0.973**</b>     | <b>0.839**</b>     | <b>0.932*</b>       | 1.477                   | 1.334 | 3.054  | 1.001             | <b>0.983</b>       | <b>0.778*</b>      | <b>0.876**</b>      |
| T-RMSFE  | 1.190                   | 1.102 | 1.418          | <b>0.994</b>      | <b>0.970</b>       | <b>0.876</b>       | <b>0.910</b>        | 1.890                   | 1.230 | 2.151  | <b>0.995</b>      | <b>0.972</b>       | <b>0.735</b>       | <b>0.846</b>        |

*Note:* This table summarizes Relative Mean Squared Forecast Errors and Trace Root Mean Squared Forecast Error (TRMPSE) relative to the Random Walk obtained by using each of the competing models, for the horizons 1-week, 1-, 3-, and 6-month-ahead. The first column in the table reports the value of (T) RMSFE (expressed in basis points) for the Random Walk model (RW), while all other lines reports statistics relative to the RW. The following model abbreviations are used in the table: RW stands for the Random Walk, (V)AR for the first-order (Vector) Autoregressive Model, DNS for the one-step dynamic Nelson–Siegel model with a (V)AR specification for the factors, AFNS refers to the one-step arbitrage-free Nelson Siegel model with a (V)AR specification for the factors. Numbers smaller than one (shown in bold) indicate that models outperform the random walk, whereas numbers larger than one indicate underperformance. The gray box indicates outperformance in the maturity. The stars on the right of the cell entries signal the level at which the [Giacomini and White \(2006\)](#) test rejects the null of equal forecasting accuracy (\*, \*\*, and \*\*\* mean respectively rejection at 1%, 5%, and 10% level).

is difficult (see, for example, [Duffee, 2002](#); [Ang and Piazzesi, 2003](#); [Diebold and Li, 2006](#); [Hoerdahl et al., 2006](#); [Moench, 2008](#)).

**3.2.1.2. Univariate autoregressive model.** Another important benchmark is the AR(1) model, which contains the RW as a special case. Thus, yields at each maturity are also predicted using a first-order univariate autoregressive model, that is estimated on the available data for that maturity:

$$y_t(\tau) = \alpha + \beta y_{t-1}(\tau) + \varepsilon_t, \quad (29)$$

for maturity  $\tau$ . The 1-step ahead forecast is produced as  $\hat{y}_{t+1}(\tau) = \hat{\alpha} + \hat{\beta} y_{t-1}(\tau)$  the forecasts for  $h$ -step ahead horizon are obtained as:

$$\hat{y}_{t+h|t}(\tau) = (1 + \hat{\beta} + \hat{\beta}^2 + \dots + \hat{\beta}^{h-1})\hat{\alpha} + \hat{\beta}^h y_t(\tau).$$

**3.2.1.3. Vector autoregressive model.** To capture some dependence between yields of different maturities, a first-order unrestricted vector autoregressive model, VAR, for yield levels is estimated. Specifically, VAR models allow the usage of the history of other maturities as additional information on top of any maturity's own history. The regression model is:

$$\mathbf{y}_t = \mathbf{A} + \mathbf{B}\mathbf{y}_{t-1} + \varepsilon_t, \quad (30)$$

where  $\mathbf{y}_t = (y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N))'$ . The 1-step ahead forecast is produced as  $\hat{\mathbf{y}}_t = \hat{\mathbf{A}} + \hat{\mathbf{B}}\mathbf{y}_{t-1}$ , while the  $h$ -step ahead forecasts are obtained as:

$$\hat{\mathbf{y}}_{t+h|t} = (I + \hat{\mathbf{B}} + \hat{\mathbf{B}}^2 + \dots + \hat{\mathbf{B}}^{h-1})\hat{\mathbf{A}} + \hat{\mathbf{B}}^h \mathbf{y}_t. \quad (31)$$

As argued by [de Pooter et al. \(2010\)](#), a well-known drawback of using an unrestricted VAR model for yields is the large number of parameters that need to be estimated. Since we want to construct forecasts for thirteen maturities, this results in a substantial number of parameters that needs to be estimated.

### 3.3. Results

[Tables 2 and 3](#) report the RMSFE and TRMSFE for each of the forecasting models, and for each of the forecasting horizons. The first column in each panel of each Table reports the value of RMSFE and TRMSFE (expressed in basis points) for the random walk model (RW), while all other columns report ratios of these statistics relative to RW. The following model abbreviations are used in [Tables 2 and 3](#): AR(1) for the first-order univariate autoregressive model, VAR(1) for the first-order vector autoregressive model, DNS<sub>AR</sub> for the dynamic Nelson–Siegel model with a (V)AR specification for the factors, and AFNS for the dynamic Arbitrage-Free Nelson Siegel model with a (V)AR specification for the factors.

For each forecasting horizon considered, we highlight the most accurate model, in terms of RMSFE, for each of the 13 maturities analyzed. Bold values indicate relative differences below one, which means that a particular model outperforms random walk. One, two and three stars denote rejection of the null hypothesis of equal forecasting ability according to the GW test at 10%, 5% and 1% level, respectively.

Empirical results show the well-documented good forecasting performance of the random walk model for short forecasting horizons (see, among others, [Diebold and Li, 2006](#); [de Pooter et al., 2010](#); [Nyholm and Vidova-Koleva, 2012](#); [Xiang and Zhu, 2013](#)). This forecasting dominance is due to the fact that, in the short-run, the near-unit-root behavior of yields tend to dominate the information brought by the structure of the yield curve model. Indeed, when looking at the RMSE for the 1-week-ahead forecasting horizon, we observe that in very few instances the candidate models are able to outperform the RW specification. Moreover, all candidate models delivered higher T-RMSFE in comparison to RW. However, already for the 1-month-ahead forecasting horizon it is possible to observe some improvements with respect to the RW, and the DNS VAR model is a clear winner, beating the RW for all maturities but the 3-month.

When longer forecasting horizons are considered (3- to 6-month-ahead), the AFNS model with independent factors (AFNS-AR) shows the lowest RMSFE for maturities higher than 9 months, whereas the AFNS model with correlated factors exhibit lowest RMSFE for the two shortest maturities (3- and 6-month). However, DNS-VAR forecasts are also significantly more accurate than those from the random-walk, except for the 3 months yields. Overall, we find that

Table 4

Trace relative mean squared forecast errors, for out-of-sample forecasts.

| Model                    | RW        | AR(1)     | VAR(1)    | DNS-AR    | DNS-VAR   | AFNS-AR   | AFNS-VAR |
|--------------------------|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| <i>One-week ahead</i>    |           |           |           |           |           |           |          |
| RW                       | 0.000     |           |           |           |           |           |          |
| AR(1)                    | 0.001     | 0.000     |           |           |           |           |          |
| VAR(1)                   | 0.007**   | 0.007**   | 0.000     |           |           |           |          |
| DNS-AR                   | 0.005***  | 0.005***  | −0.002    | 0.000     |           |           |          |
| DNS-VAR                  | 0.008**   | 0.008**   | 0.001     | 0.003     | 0.000     |           |          |
| AFNS-AR                  | 0.009**   | 0.008**   | 0.001     | 0.003     | 0.000     | 0.000     |          |
| AFNS-VAR                 | 0.003     | 0.003     | −0.004    | −0.002    | −0.005*** | −0.005*** | 0.000    |
| <i>One-month ahead</i>   |           |           |           |           |           |           |          |
| RW                       | 0.000     |           |           |           |           |           |          |
| AR(1)                    | 0.009     | 0.000     |           |           |           |           |          |
| VAR(1)                   | 0.026***  | 0.017     | 0.000     |           |           |           |          |
| DNS-AR                   | 0.001     | −0.008    | −0.025*** | 0.000     |           |           |          |
| DNS-VAR                  | −0.012*** | −0.021*** | −0.038**  | −0.013    | 0.000     |           |          |
| AFNS-AR                  | 0.058*    | 0.048**   | 0.032**   | 0.057*    | 0.069*    | 0.000     |          |
| AFNS-VAR                 | 0.004     | −0.005    | −0.022    | 0.003     | 0.016     | −0.053**  | 0.000    |
| <i>Three-month ahead</i> |           |           |           |           |           |           |          |
| RW                       | 0.000     |           |           |           |           |           |          |
| AR(1)                    | 0.302**   | 0.000     |           |           |           |           |          |
| VAR(1)                   | 1.431*    | 1.129*    | 0.000     |           |           |           |          |
| DNS-AR                   | −0.017    | −0.319**  | −1.448*   | 0.000     |           |           |          |
| DNS-VAR                  | −0.083    | −0.385*   | −1.514*   | −0.066    | 0.000     |           |          |
| AFNS-AR                  | −0.329*   | −0.631*   | −1.760*   | −0.312**  | −0.246*** | 0.000     |          |
| AFNS-VAR                 | −0.243*** | −0.545**  | −1.674*   | −0.226*** | −0.160*** | 0.086     | 0.000    |
| <i>Six-month ahead</i>   |           |           |           |           |           |           |          |
| RW                       | 0.000     |           |           |           |           |           |          |
| AR(1)                    | 1.830**   | 0.000     |           |           |           |           |          |
| VAR(1)                   | 12.978*   | 11.147*   | 0.000     |           |           |           |          |
| DNS-AR                   | −0.038    | −1.869**  | −13.016*  | 0.000     |           |           |          |
| DNS-VAR                  | −0.198    | −2.029*   | −13.176*  | −0.160    | 0.000     |           |          |
| AFNS-AR                  | −1.646**  | −3.476*   | −14.624*  | −1.608**  | −1.448**  | 0.000     |          |
| AFNS-VAR                 | −1.019**  | −2.850*   | −13.997*  | −0.981*** | −0.821*** | 0.627***  | 0.000    |

Note: This table presents the difference of the trace mean square forecast errors (T-RMSFE) obtained by using each of the competing models, for the horizons 1-week, 1-, 3-, and 6-month-ahead, comparing the methods by pairs. Negative values indicate superiority of the forecast for the row method of the pair. The stars on the right of the cell entries signal the level at which the [Giacomini and White \(2006\)](#) test rejects the null of equal forecasting accuracy (\*, \*\*, and \*\*\* mean respectively rejection at 1%, 5%, and 10% level, respectively).

the AFNS-AR model shows the lowest RMSFE for 3- and 6-month ahead forecast and it is the only one that exhibit significant RMSFE superiority relative to RW for all maturities considered. Nevertheless, the performance of both DNS and AFNS models is clearly superior to the performance of the benchmarks considered.

Note, however, that GW tests are all performed with respect to the RW model, which does not allow us to statistically compare the results of DNS and AFNS models directly. To gain more insight in the relative performance between of no-arbitrage restrictions, we calculate the difference of trace mean square forecast errors (T-RMSFE) for all pairs of competing models and assess the statistical significance of these forecasting differences through the Giacomini-White test.<sup>7</sup> Results are shown in [Table 4](#). Negative entries indicate the superiority of the first model of the pair (in the line). [Table 4](#) shows that no-arbitrage restrictions seems to improve forecasts. Particularly, the AFNS-VAR model is significantly more accurate than its unrestricted counterpart (the DNS-VAR model) at the 1% level for the one-week, three-months, and six-months ahead horizons. The AFDNS-AR model is also superior to the DNS-AR model for the

<sup>7</sup> We do not perform pair-wise comparison for all maturities because this would require 14 additional tables (one for each maturity).



three-months, and six-months ahead horizons. However, the DNS-AR model significantly outperforms the AFDNS for the one-month ahead horizon.

Our article is closely related to those of [Vicente and Tabak \(2008\)](#), [Caldeira et al. \(2010\)](#), [Araújo and Cajueiro \(2013\)](#), and [Caldeira et al. \(2016\)](#), that have investigated the forecasting ability of the dynamic Nelson–Siegel models for the Brazilian yield curve. Indeed, our results are consistent with those reported for them. [Vicente and Tabak \(2008\)](#) find that DNS model provides superior forecasts relative to random walk, especially at longer time horizons for short-term interest rates. [Araújo and Cajueiro \(2013\)](#), [Caldeira et al. \(2016\)](#) reveal that it is not possible to determine an individual model that consistently produces superior forecasts for all maturities and all forecasts horizons. However, similar to us, they find that the DNS model is competitive in the out-of-sample forecasting of bond yields, mainly when we consider 1- and 3-month ahead horizon. Moreover, our empirical analysis suggests that the arbitrage-free dynamic Nelson–Siegel term structure model proposed by [Christensen et al. \(2011\)](#) is helpful for forecasting the Brazilian yield curve, especially for longer horizons. For DNS and AFNS models, the general pattern can be summarized as follows. Forecasts are more accurate if the horizon is longer and if the maturity is shorter for the AFNS, and maturity longer for DNS, as can be seen from [Tables 2 and 3](#).

In summary, our empirical evidence suggests that when a short forecasting horizon is considered (e.g. 1-week-ahead) the differences in forecasting performance among the candidate models is rather inconclusive since in very few instances the candidate models outperform the baseline RW model. However, when longer forecasting horizons are considered, both the DNS and AFNS models deliver good forecasts. The most important message from our empirical test is that the imposition of no-arbitrage is indeed helpful for longer forecasting horizons. These results corroborate the evidence reported in [Christensen et al. \(2011\)](#) for US Treasuries data, as they find that the AFNS model with independent factors outperform the DNS specifications for the 6-month- and 12-month-ahead forecast horizons.

#### 4. Concluding remarks

The dynamic version of the Nelson–Siegel model has been shown in the literature to be remarkably well suited both to fit and to forecast the term structure of interest rates. More recently, [Christensen et al. \(2011\)](#) have developed an arbitrage-free version of this model in order to bring theoretical rigor to a empirically successful model. In this paper, we have analyzed the forecasting power of no-arbitrage restrictions on the dynamic Nelson–Siegel model in order to determine the empirical relevance of this new theoretical restriction.

Different specification for the factor dynamics are allowed in both versions of the DNS model. Their forecasts are compared on a yield to yield basis using the root mean squared forecast error, and their overall performance is measured via the trace root mean squared error. The empirical results corroborates the well-known phenomenon of the good forecasting performance of the random walk model for the shorter forecasting horizon. For the 3- and 6-months forecasting horizons, the Nelson–Siegel model and its no-arbitrage counterpart outperform the random walk and other competitors. Specifically, for 3- and 6-months ahead forecasts, the Giacomini-White test indicate superiority of the AFNS model relative to RW for all considered maturities. Moreover, a direct comparison of T-RMSE from AFNS and DNS models show that arbitrage restrictions significantly improve longer horizons forecasts. Thus, we provide evidence that the imposition of no-arbitrage is indeed helpful for longer forecasting horizons. However, arbitrage-free restrictions do not seem to improve short horizon forecasts.

#### Acknowledgments

João F. Caldeira gratefully acknowledges support provided by CNPq under grants 309899/2015-0 and 485561/2013-1. Guilherme V. Moura gratefully acknowledges support provided by CNPq under grants 475230/2012-4, and 308346/2013-0. André A.P. Santos gratefully acknowledges supports provided by CNPq under grants 408723/2013-0, 406493/2012-0, and 307320/2013-8.

#### References

- Almeida, C., Gomes, R., Leite, A., Simonsen, A., Vicente, J., 2009. Does curvature enhance forecasting? *Int. J. Theor. Appl. Finance* 12 (08), 1171–1196.
- Almeida, C., Vicente, J., 2008. The role of no-arbitrage on forecasting: lessons from a parametric term structure model. *J. Bank. Finance* 32 (12), 2695–2705.

- Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *J. Monet. Econ.* 50 (4), 745–787.
- Araújo, R.C., Cajueiro, D.O., 2013. Combining term structure of interest rate forecasts: the Brazilian case. *Economía* 14 (2), 102–121.
- Björk, T., Christensen, B.J., 1999. Interest rate dynamics and consistent forward rate curves. *Math. Finance* 9 (4), 323–348.
- Caldeira, J.F., Moura, G.V., Portugal, M.S., 2010. Efficient yield curve estimation and forecasting in Brazil. *Economía* 11 (1), 27–51.
- Caldeira, J.F., Moura, G.V., Santos, A.P., 2016. Predicting the yield curve using forecast combinations. *Comput. Stat. Data Anal.* 100 (August), 79–98.
- Carriero, A., Giacomini, R., 2011. How useful are no-arbitrage restrictions for forecasting the term structure of interest rates? *J. Econom.* 164 (1), 21–34.
- Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D., 2009. An arbitrage-free generalized Nelson–Siegel term structure model. *Econom. J.* 119, 33–64.
- Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D., 2011. The affine arbitrage-free class of Nelson–Siegel term structure models. *J. Econom.* 164 (1), 4–20.
- Coroneo, L., Nyholm, K., Vidova-Koleva, R., 2011. How arbitrage-free is the Nelson–Siegel model? *J. Empir. Finance* 18 (3), 393–407.
- de Pooter, M., Ravazzolo, F., van Dijk, D., 2010, March. Term structure forecasting using macro factors and forecast combination. Working Paper 2010/01. Norges Bank.
- de Rezende, R.B., Ferreira, M.S., 2013. Modeling and forecasting the yield curve by an extended Nelson–Siegel class of models: a quantile autoregression approach. *J. Forecast.* 32 (2), 111–123.
- Diebold, F., Li, C., 2006. Forecasting the term structure of government bond yields. *J. Econom.* 130 (2), 337–364.
- Diebold, F.X., Rudebusch, G.D., 2013. *The Dynamic Nelson–Siegel Approach to Yield Curve Modeling and Forecasting*. Princeton University Press.
- Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. *J. Finance* 57 (1), 405–443.
- Duffee, G.R., 2011. Forecasting with the term structure: The role of no-arbitrage restrictions. Working paper. Johns Hopkins University, Department of Economics.
- Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. *Math. Finance* 6 (4), 379–406.
- Favero, C.A., Niu, L., Sala, L., 2012. Term structure forecasting: no-arbitrage restrictions versus large information set. *J. Forecast.* 31 (2), 124–156.
- Filipovic, D., 1999. A note on the Nelson–Siegel family. *Math. Finance* 9 (4), 349–359.
- Geweke, J., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. *Rev. Financ. Stud.* 9 (2), 557–587.
- Giacomini, R., White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74 (6), 1545–1578.
- Gimeno, R., Marqués, J.M., 2009. Extraction of financial market expectations about inflation and interest rates from a liquid market, Working Paper 0906. Banco de España.
- Golub, G.H., Van Loan, C.F., 1996. *Matrix Computations*, 3rd ed. Johns Hopkins University Press.
- Hoerdahl, P., Tristani, O., Vestin, D., 2006. A joint econometric model of macroeconomic and term-structure dynamics. *J. Econom.* 131 (1–2), 405–444.
- Joslin, S., Singleton, K.J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. *Rev. Financ. Stud.* 24 (3), 926–970.
- Jungbacker, B., Koopman, S.J., 2014. Likelihood-based dynamic factor analysis for measurement and forecasting. *Econom. J.* 118, C1–C21.
- Lang, S., 1987. *Linear Algebra*, 3rd ed. Springer.
- Marcellino, M., Stock, J., Watson, M., 2005. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *J. Econom.* 135 (1–2), 499–526.
- Moench, E., 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented VAR approach. *J. Econom.* 146 (1), 26–43.
- Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. *J. Bus.* 60 (4), 473–489.
- Nyholm, K., Vidova-Koleva, R., 2012. Nelson–Siegel, affine and quadratic yield curve specifications: which one is better at forecasting? *J. Forecast.* 31 (6), 540–564.
- Pesaran, M.H., Pick, A., Timmermann, A., 2011. Variable selection, estimation and inference for multi-period forecasting problems. *J. Econom.* 164 (1), 173–187.
- Svensson, L.E.O., 1994, September. Estimating and interpreting forward interest rates: Sweden 1992–1994. IMF Working Papers 94/114. International Monetary Fund.
- Vereda, L., Lopes, H., Fukuda, R., 2008. Estimating VAR models for the term structure of interest rates. *Insur.: Math. Econ.* 42 (2), 548–559.
- Vicente, J., Tabak, B.M., 2008. Forecasting bond yields in the Brazilian fixed income market. *Int. J. Forecast.* 24 (3), 490–497.
- Williams, D., 1997. *Probability with Martingales*. Cambridge University Press.
- Xiang, J., Zhu, X., 2013. A regime-switching Nelson–Siegel term structure model and interest rate forecasts. *J. Financ. Econom.* 11 (3), 522–555.